HL7 Clinical Genomics – From Research to Healthcare

NCI/NCRI Joint Conference
June 13-14, 2011

Amnon Shabo (Shvo), PhD

HL7 Clinical Genomics WG
Co-chair and Modeling Facilitator

HL7 Structured Documents WG
CDA R2 Co-editor
CCD Implementation Guide Co-editor

Agenda

- HL7 Clinical Genomics
- New Specifications
- Experimental Implementations

The Mission of HL7 Clinical Genomics Work Group

- The HL7 Clinical Genomics Work Group (CGWG) supports the HL7 mission to create and promote its standards by enabling the communication between interested parties of clinical and genomic data related to an individual. Part of the CGWG efforts is the personalization of genomic data – the so-call 'omics differences in an individual's genome and its association with relevant phenotypic and clinical information. Associations to interventive/expected phenotypes will be modeled as knowledge that can be utilized to transform an individual's data into meaningful information.

- CGWG will facilitate the development of common standards for clinical research information management across a variety of organizations -- including national and international government agencies and regulatory bodies, private efforts, and sponsored research -- and thus the availability of safe and effective therapies by improving the processes and efficiencies associated with regulated clinical research.

- CGWG will strive to achieve common semantics between clinical and research environments. Consequently each standardization effort is to start with definitions that later on can be refined to specific realms.

Healthcare

Research

Common Semantics

Overview of Activities

Three Tracks:

v3:

- Family History (Pedigree) Topic
- Genetic Variations Topic
- Gene Expression Topic
- CMETs defined by the Domain

v2:

v2 Implementation Guides

* The IG "Genetic Test Result Reporting to EHR" is modeled after the HL7 Version 2.5.1 Implementation Guide: Orders And Observations; Interoperable Laboratory Result Reporting To EHR (US Realm), Release 1

CDA:

- A CDA Implementation Guide for Genetic Testing Reports

Common:

- Domain Analysis Models for the various topics
- A Domain Information Model (v3) describing the common semantics
- Semantic a*lignment among the various specs*

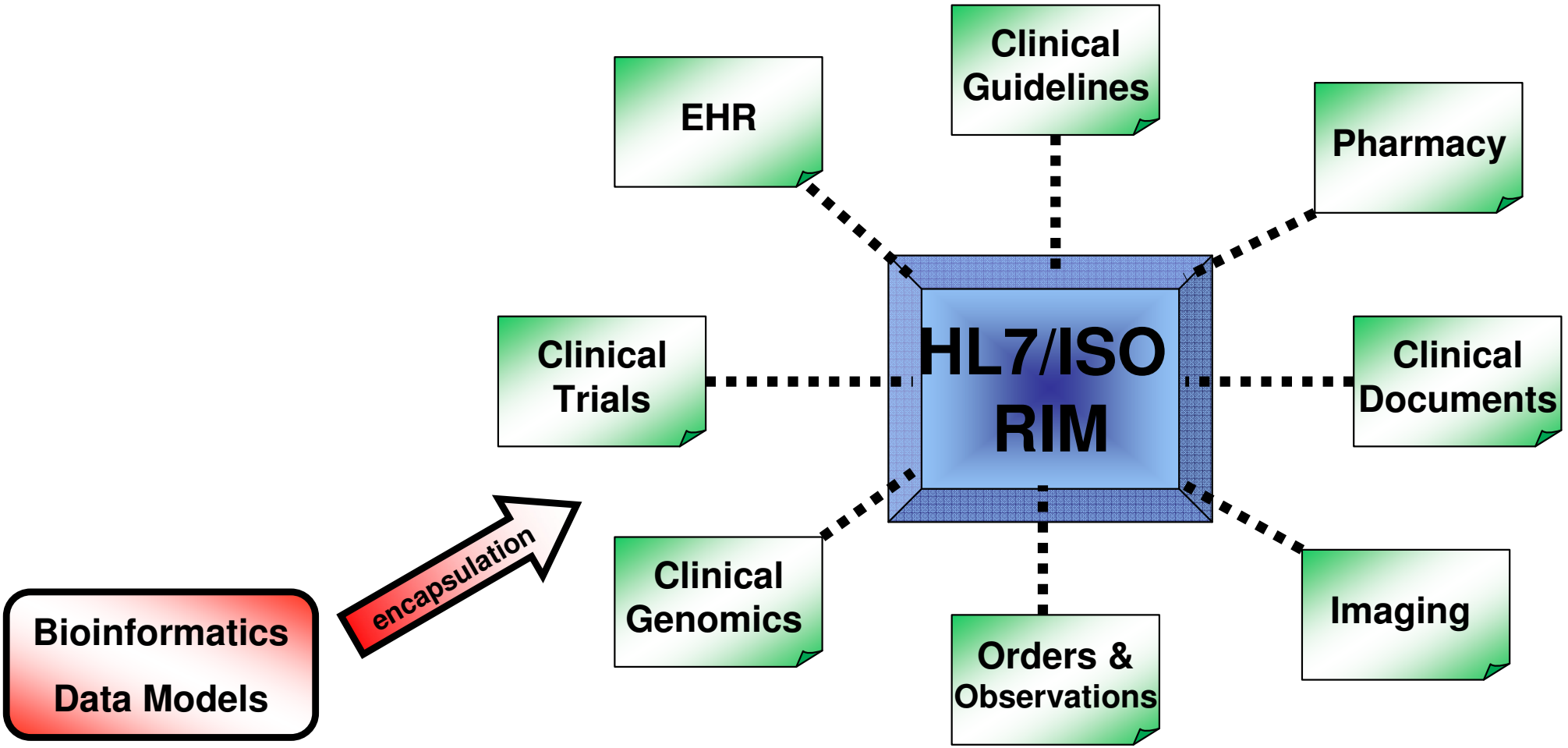
Normative
DSTU
Informative

Main Principles

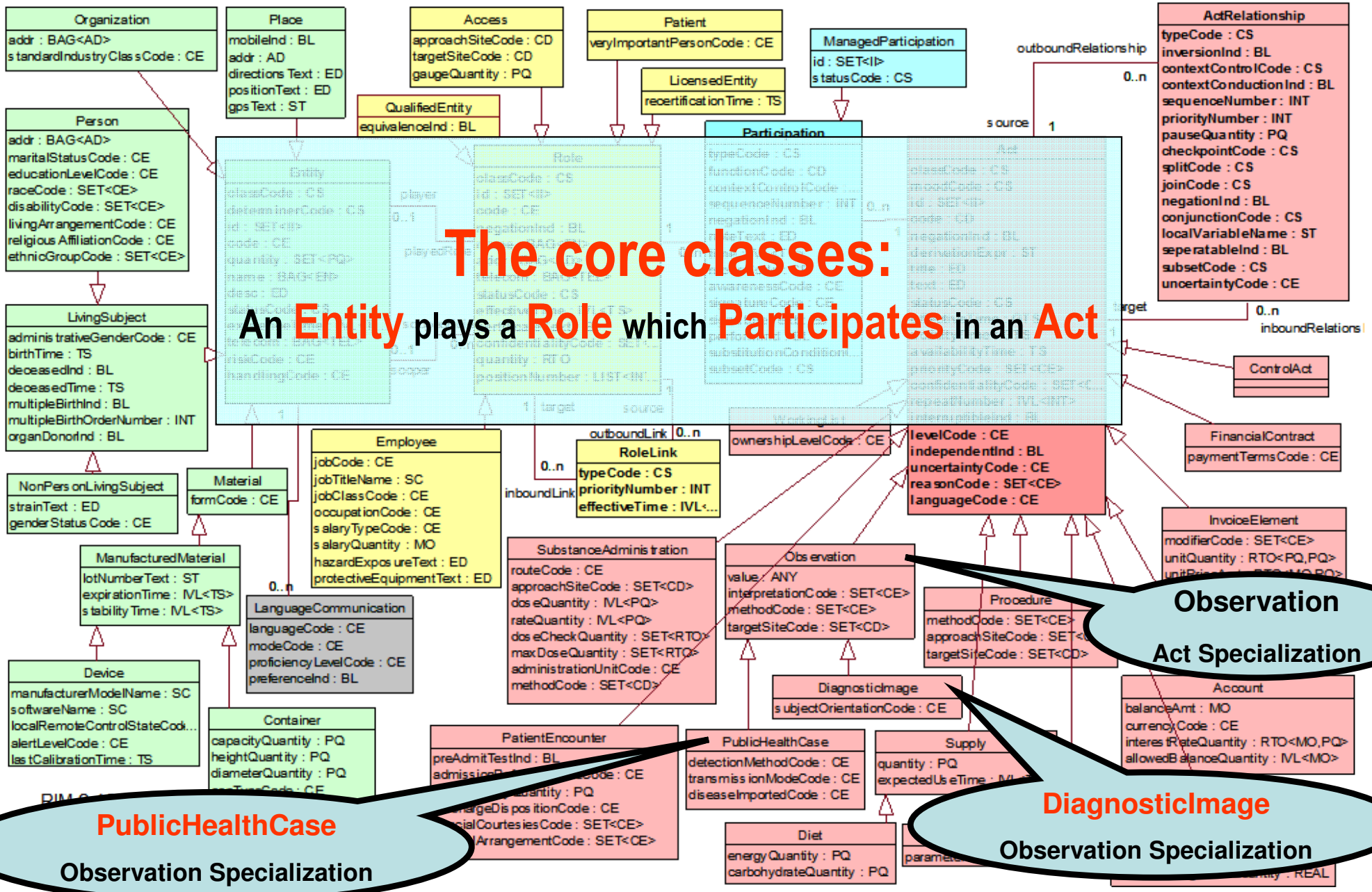
- **Underlying Model**
  - HL7/ISO Reference Information Model
- **Clinical Genomics Statement**
  - Standard grammar of genotype-phenotype associations
- **Raw genomic patient data**
  - Encapsulate and Bubble Up (through bioinformatics formats)
- **Domain Information Model**
  - The Genome model – Overarching locus and non-locus data
- **Genomics to EHR Systems**
  - CEN EHR 13606 over HL7 RIM
  - Specific Clinical Genomics Statements as:
    - DCMs or Archetypes, over the Clinical Statement model

To achieve semantic interoperability…

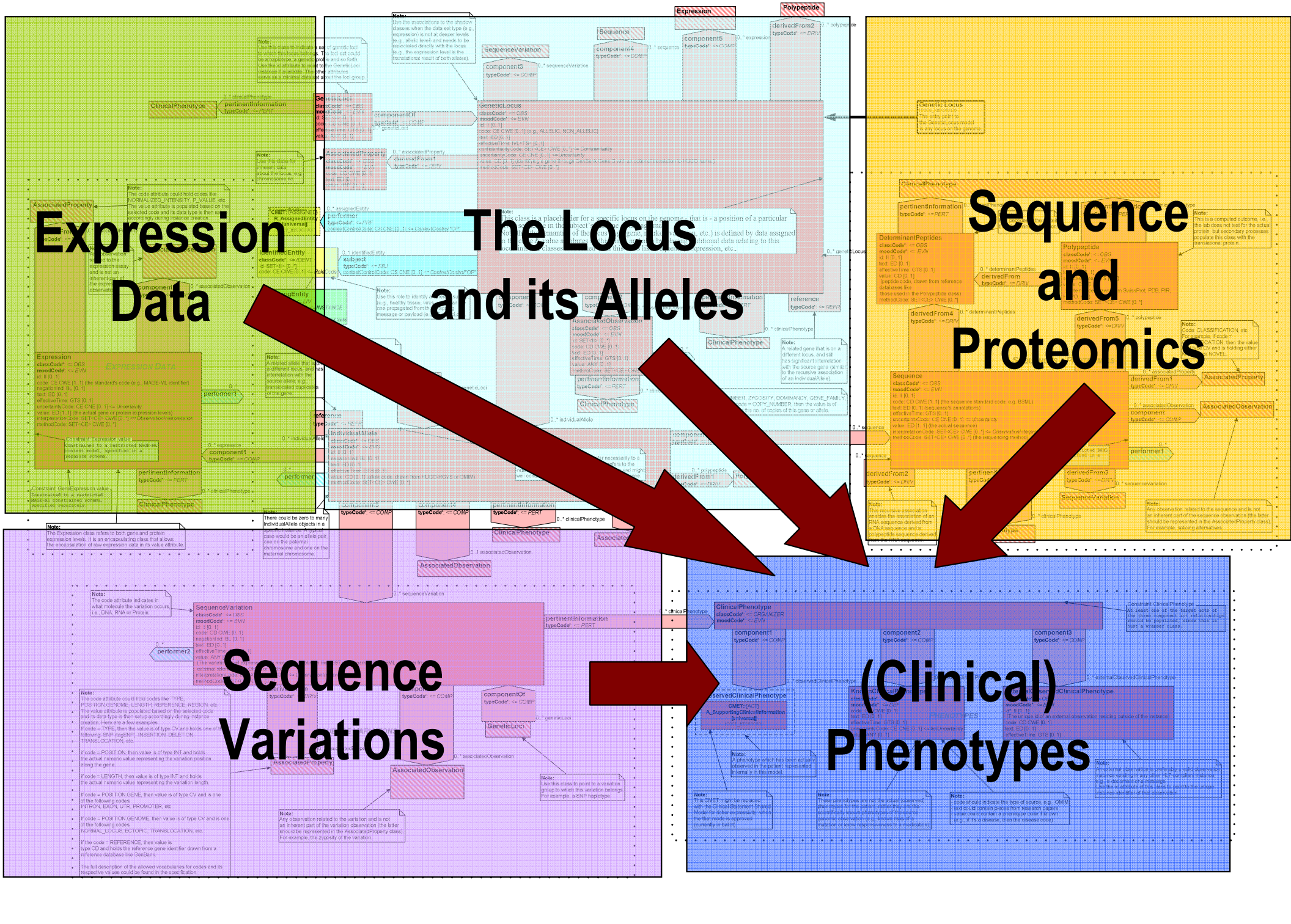
…we need standard specs derived from a reference information model:



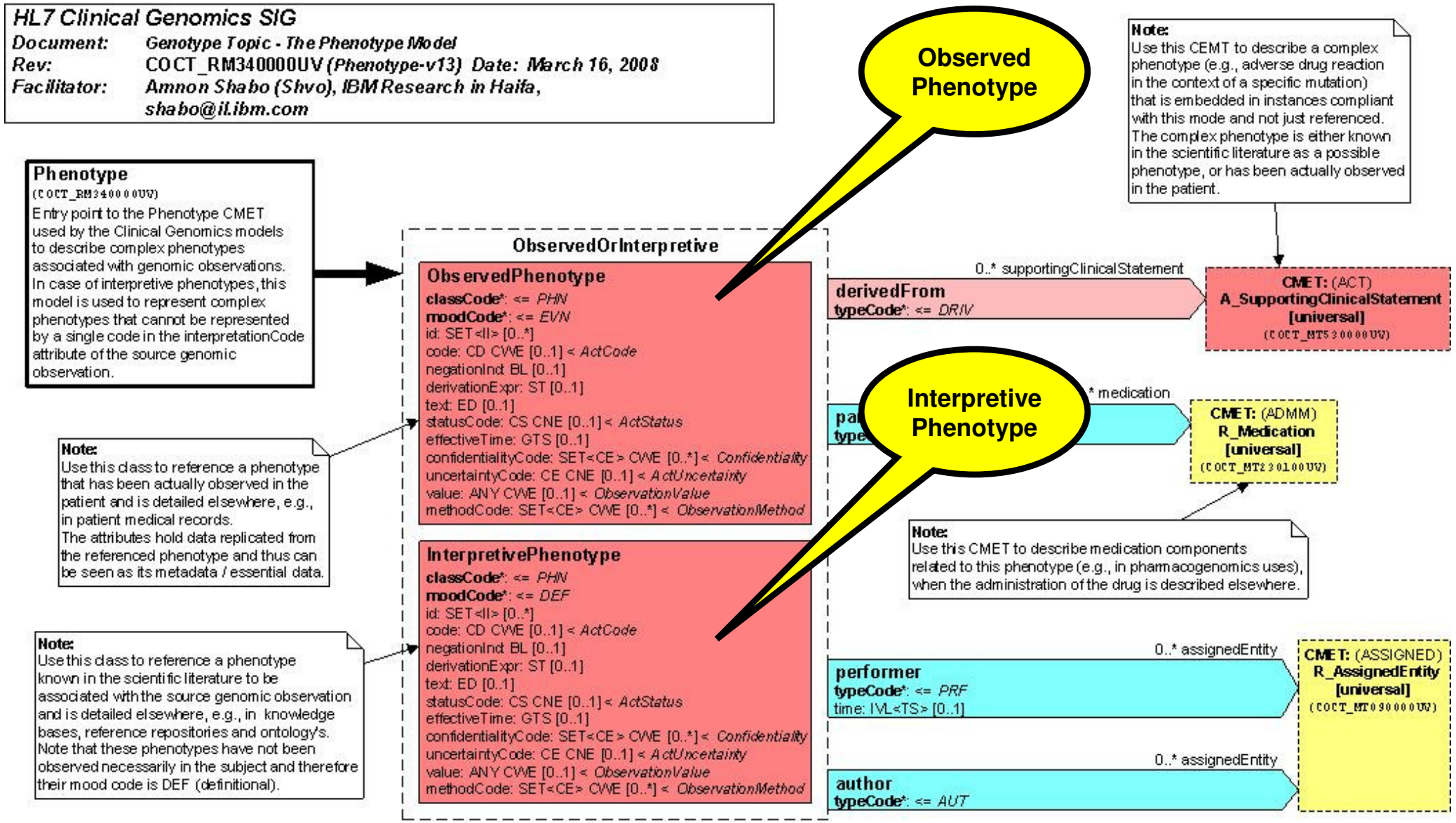
The HL7 RIM – Representing Genomics data



The DSTU GeneticLocus Model Focal Areas:

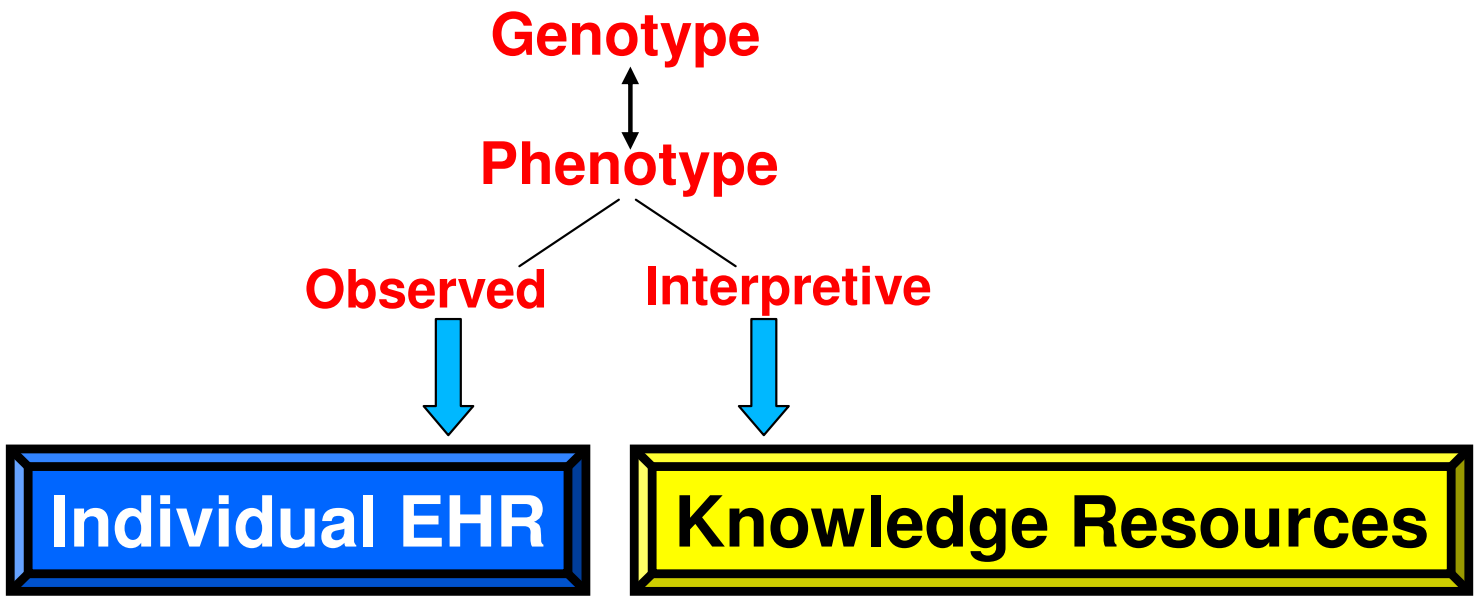


The Phenotype Model



Genotype-Phonotype Associations

- In clinical environments:
  - Observed versus interpretive phenotypes
  - Observed should reside in the EHR
  - Interpretive should be related to knowledge base

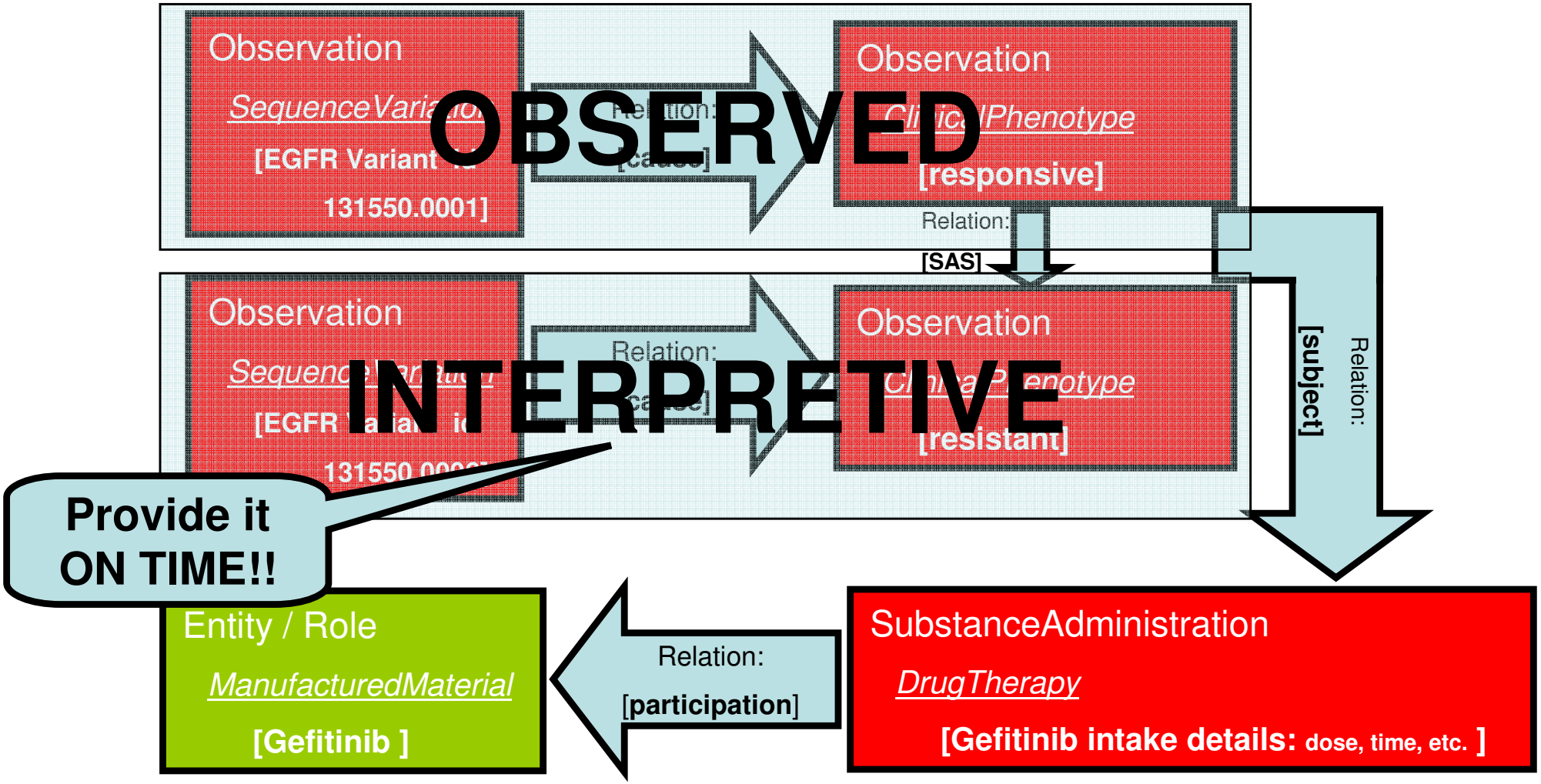


From Data to Knowledge

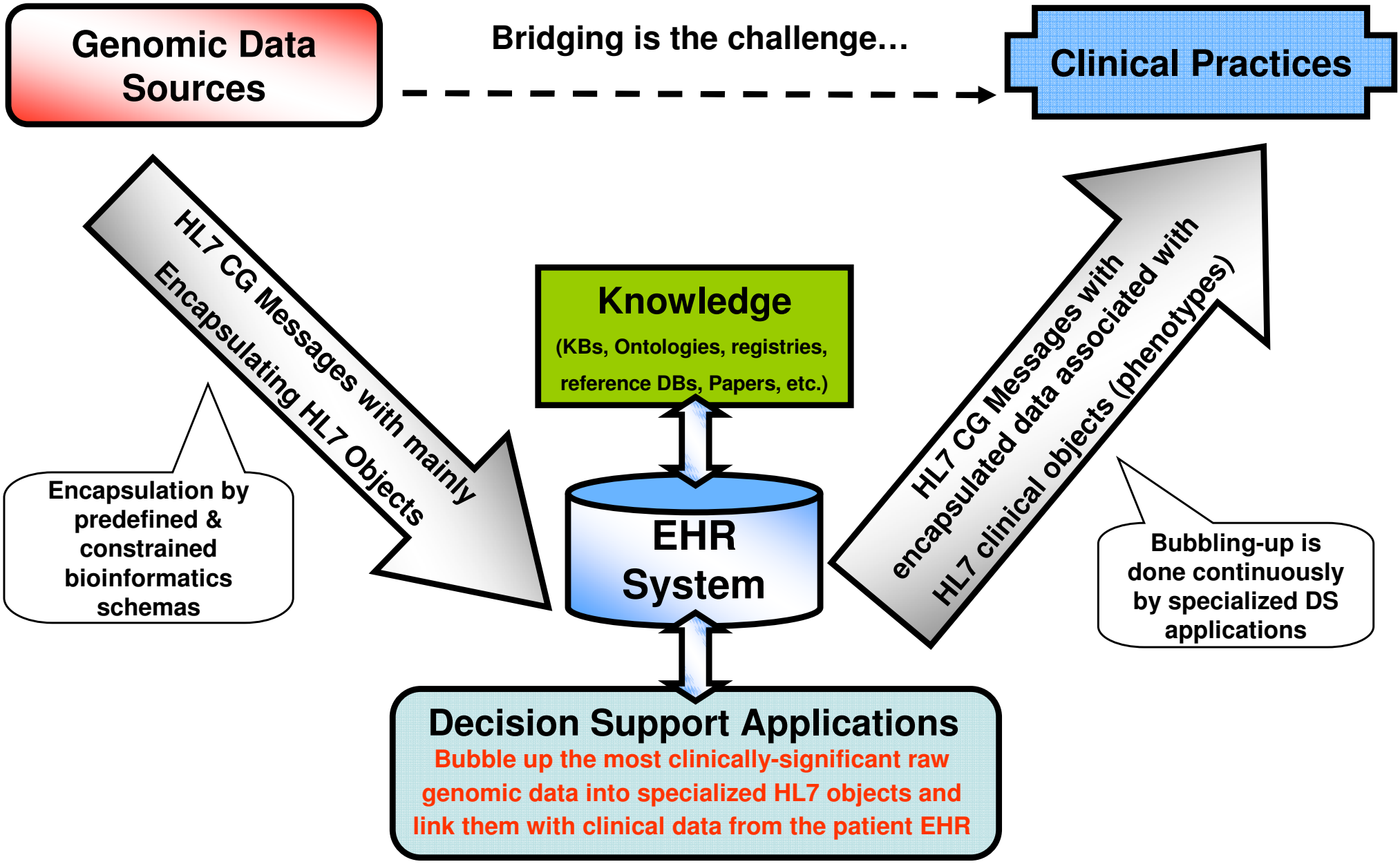
e.g., an OMIM Entry:

Despite the dramatic responses to EGFR inhibitors in patients with non-small cell lung cancer, most patients ultimately have a relapse. {12:Kobayashi et al. (2005)} reported a patient with EGFR-mutant, Gefitinib-responsive, advanced non-small cell lung cancer who had a relapse after 2 years of complete remission during treatment with Gefitinib. The DNA sequence of the EGFR gene in his tumor biopsy specimen at relapse revealed the presence of a second mutation ({131550.0006}). Structural modeling and biochemical studies showed that this second mutation led to the Gefitinib resistance.

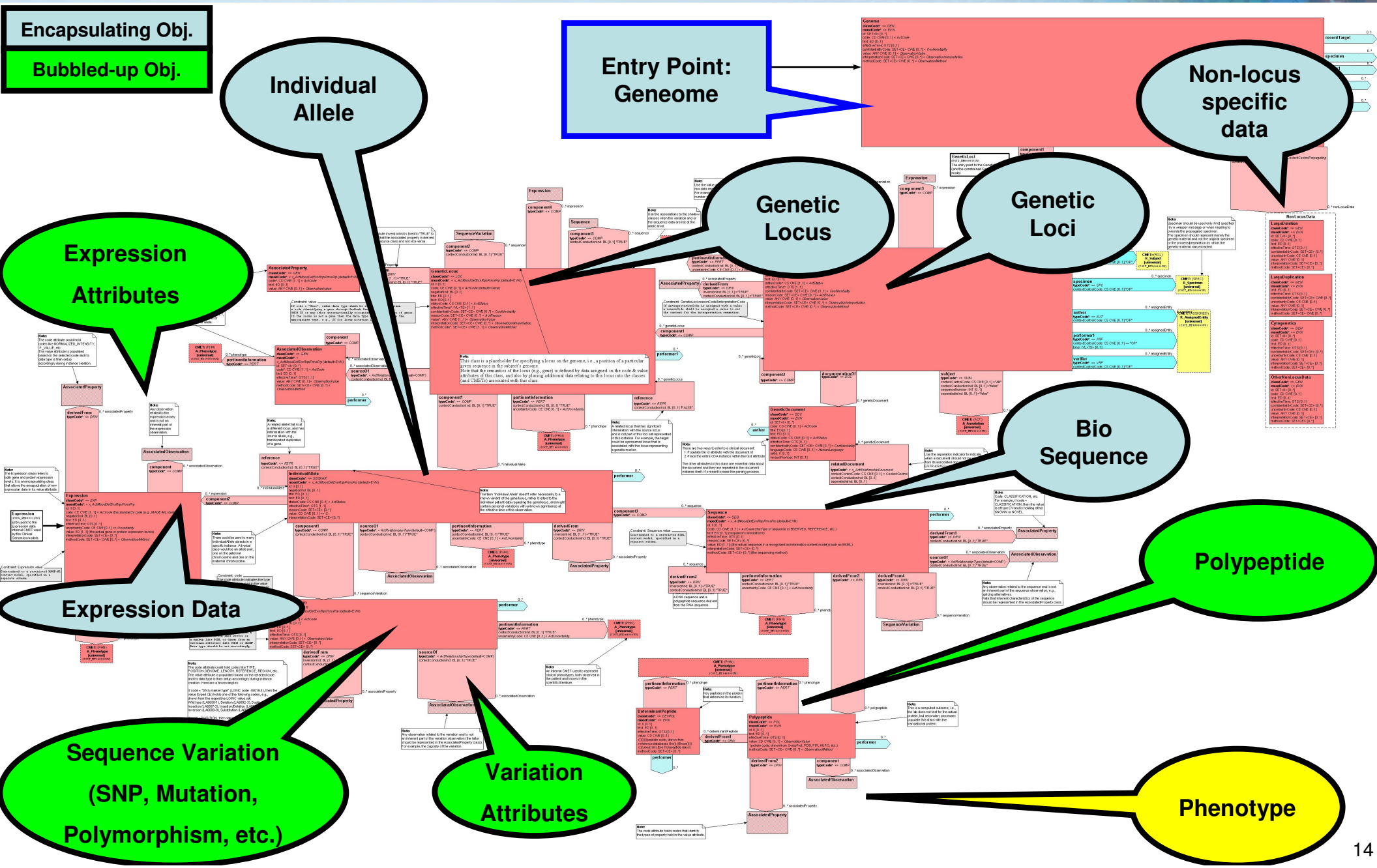
Example: Structuring OMIM Entries (cont.)



The Underlying Paradigm: Encapsulate & Bubble-up



The Domain Information Model - Genome



Example: Family History XML Encoding

Taken from a patient pedigree, the portion related to patient's daughter

```
<!-- DAUGHTER -->
- <relationshipHolder>
    <id extension="555.011" />
    <code code="DAU" />
  + <relationshipHolder>
    <!-- GENOMIC DATA -->
  - <subjectOf>
    - <clinicalGenomicChoice>
      - <clinicalGenomicChoiceGenotype>
        - <Genotype>
          - <individualAllele>
              <code code="BRCA1" codeSystem="[insert GenBank OID]"
                codeSystemName="GenBank" />
              <text>Homo sapiens breast and ovarian cancer susceptibility (BRCA1)
                complete cds.</text>
            + <AlleleSequence>
            + <SequenceVariation>
            </individualAllele>
          </Genotype>
        </clinicalGenomicChoiceGenotype>
      </clinicalGenomicChoice>
    </subjectOf>
    <!-- CLINICAL DATA -->
  + <subjectOf>
  </relationshipHolder>
  <!-- end of DAUGHTER data -->
```

Bubble up…

Point back…

To phenotype and beyond….

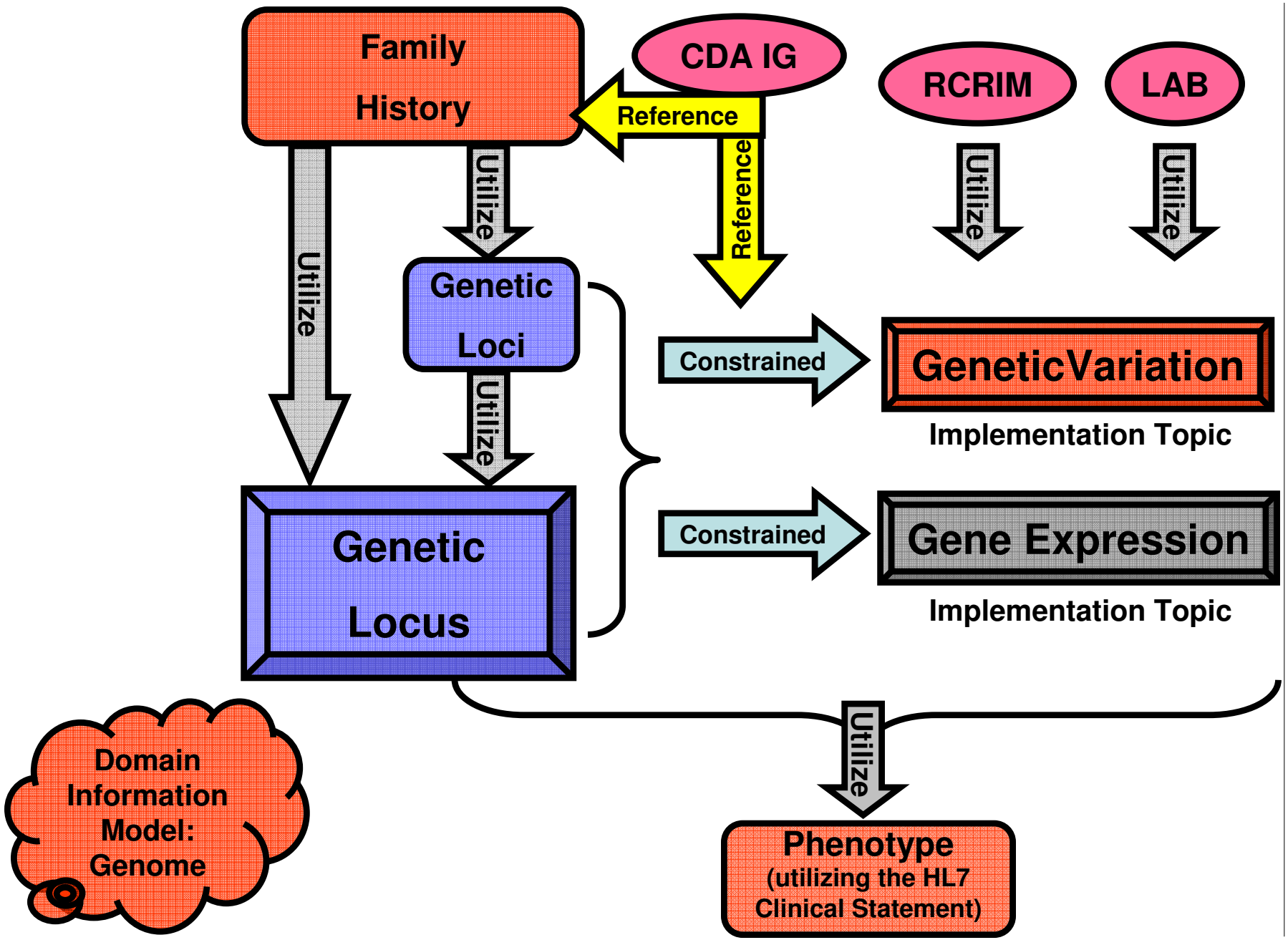
XML Fusion: Encapsulation of Raw Genomic Data

HL7 v3 XML

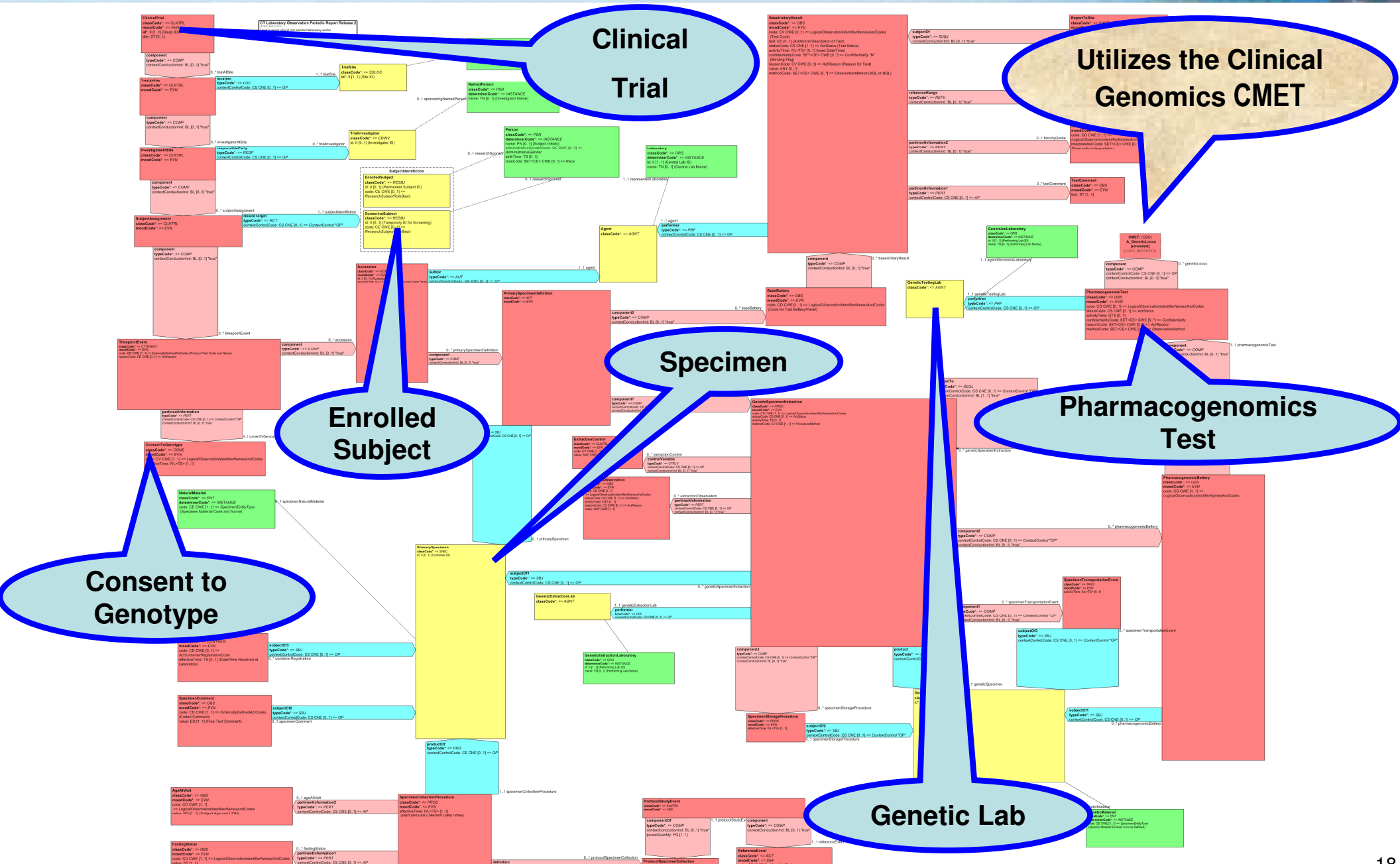
Raw genomic data represented in Bioinformatics markup

```
<subjectOf2>
  <geneticLocus>
    <component1>
      <individualAllele moodCode="EVN">
        <text>breast cancer 1, early onset</text>
        <value code="83990" displayName="BRCA1" codeSystemName="NCBI Entrez">
          <translation code="20473" displayName="BRCA1" codeSystem="HGNC"/>
        </value>
        <component2>
          <sequence moodCode="EVN">
            <code code="BSMLcon3"/>
            <value mediaType="text/xml">
              <bsml:Bsml xmlns:bsml="urn:bsml.org">
                <bsml:Definitions>
                  <bsml:Sequences>
                    <bsml:Sequence id="seq1" molecule="dna" ic-acckey="U14680 REGION: 101..199" db-source="GenBank" title="
                      BRCA1, exon 2" representation="raw" local-acckey="this could be used by the genetic lab">
                      <bsml:Seq-data>
                        GCTCCCA CTCCATGAGG TATTTCTTCA
                        CATCCGTGTC CCGGCCCGGC CGCGGGGAGC CCCGCTTCAT CGCCGTGGGC
                        TACGTGGACG ACACGCAGTT CGTGCGGTTC GACAGCGACG CCGCGAGCCA
                        GAGGATGGAG CCGCGGGCGC CGTGGATAGA GCAGGAGGGG CCGGAGTATT
                        GGGACCAGGA GACACGGAAT GTGAAGGCCC AGTCACAGAC TGACCGAGTG
                        GACCTGGGGA CCCTGCGCGG CTACTACAAC CAGAGCGAGG CCG
                      </bsml:Seq-data>
                    </bsml:Sequence>
                    <bsml:Sequence id="seq2" molecule="dna" ic-acckey="U14680 REGION: 200..253" db-source="GenBank" title="
                      BRCA1, exon 3" representation="raw" local-acckey="this could be used by the genetic lab">
                      <bsml:Seq-data>
                        GTTCTCA
                        CACCATCCAG ATAATGTATG GCTGCGACGT GGGGTCGGAC GGGCGCTTCC
                        TCCGCGGGTA CCGGCAGGAC GCCTACGACG GCAAGGATTA CATCGCCCTG
                        AACGAGGACC TGCGCTCTTG GACCGCGGCG GACATGGCGG CTCAGATCAC
                        CAAGCGCAAG TGGGAGGCGG CCCATGTGGC GGAGCAGCAG AGAGCCTACC
                        TGGATGGCAC GTGCGTGGAG TGGCTCCGCA GATACCTGGA GAACGGGAAG
                        GAGACGCTGC AGCGCACGG
                      </bsml:Seq-data>
                    </bsml:Sequence>
                  </bsml:Sequences>
```

HL7 Clinical Genomics v3 Static Models



The HL7 RCRIM CT Laboratory Model- The Pharmacogenomics Extension

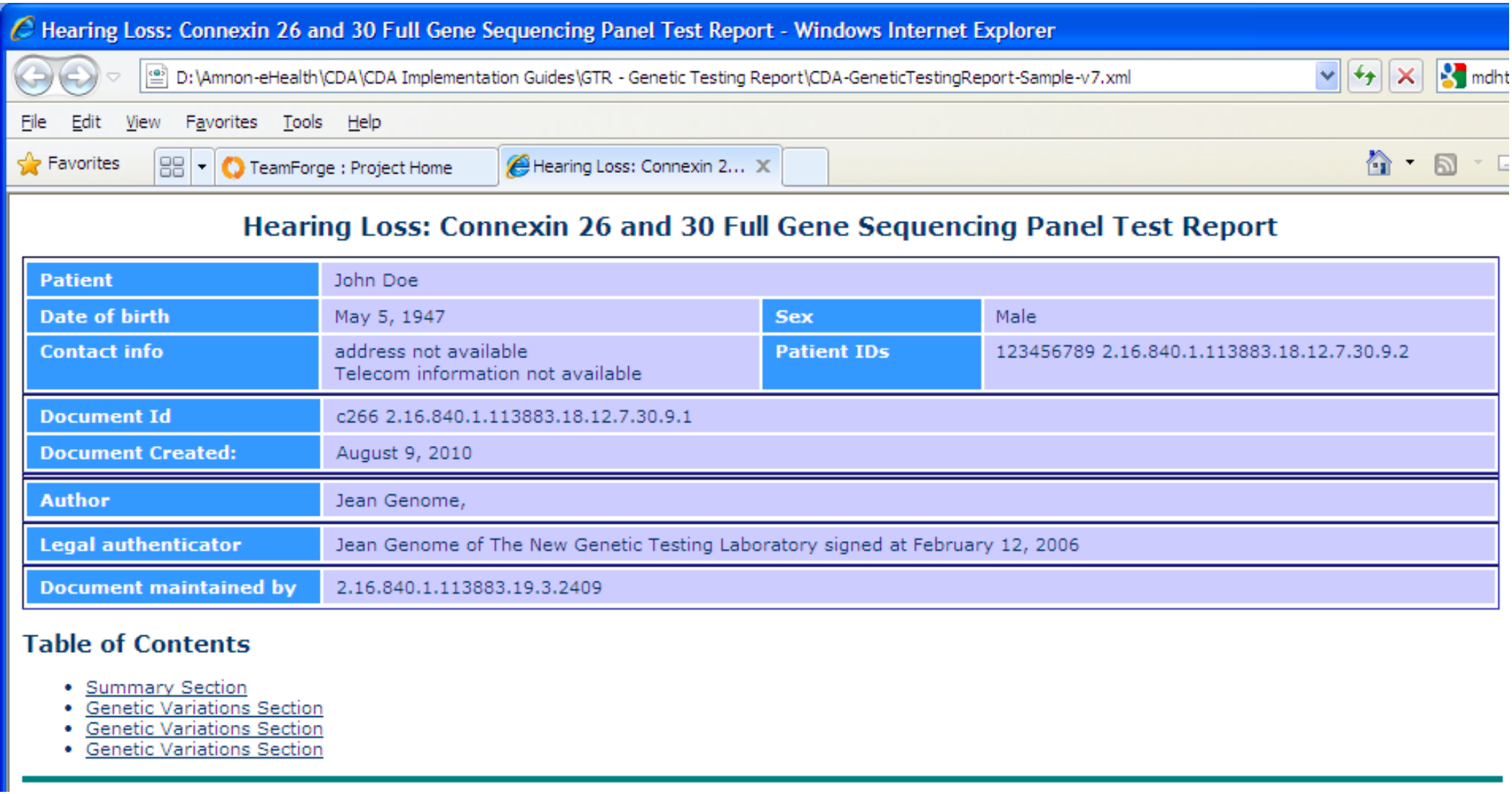


New Specifications under Ballot

CDA IG for Genetic Testing Reports

- **Scope**
  - Define a universal implementation guide for genetic testing reports that are both human readable and machine-processable
- **Design principles**
  - Follow existing report formats commonly used in healthcare & research
  - Emphasize interpretations & recommendations
  - Provide general background information on tests performed
  - Represent interpretation by utilizing patterns of 'genotype-phenotype' associations in the HL7 v3 Clinical Genomics and implement them as harmonized clinical statement entry-level templates in this IG
  - Reference HL7 Clinical Genomics instances as the place holders of raw data (personal evidences), similarly to referencing images (technically-wise)
- **CDA Template Editor:**
  - Developed using the MDHT open source tool (OHT)

GTR Rendered – The Header



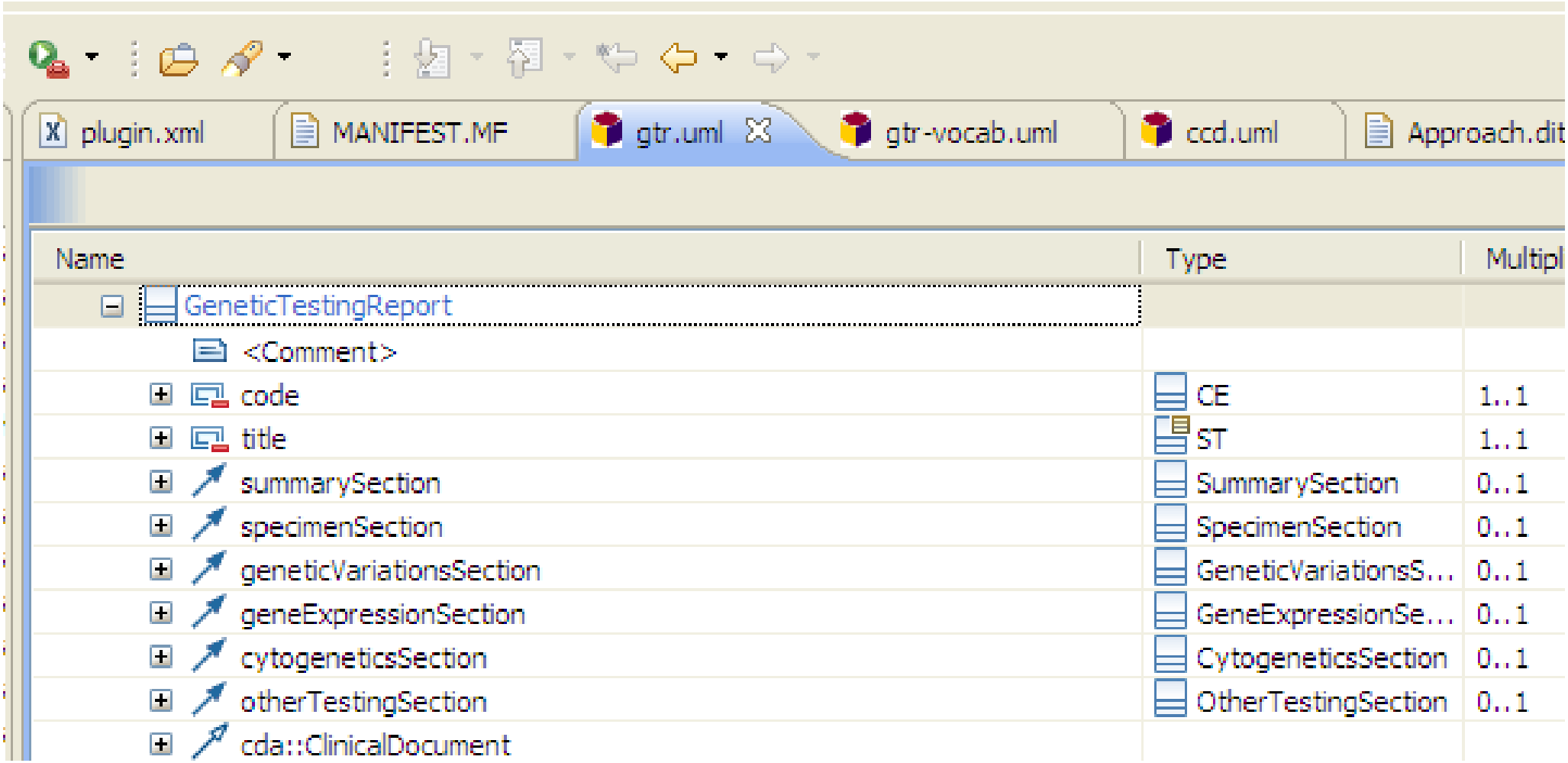
Hearing Loss: Connexin 26 and 30 Full Gene Sequencing Panel Test Report

Patient	John Doe		
Date of birth	May 5, 1947	**Sex**	Male
Contact info	address not available Telecom information not available	**Patient IDs**	123456789 2.16.840.1.113883.18.12.7.30.9.2
Document Id	c266 2.16.840.1.113883.18.12.7.30.9.1		
Document Created:	August 9, 2010		
Author	Jean Genome,		
Legal authenticator	Jean Genome of The New Genetic Testing Laboratory signed at February 12, 2006		
Document maintained by	2.16.840.1.113883.19.3.2409		

Table of Contents

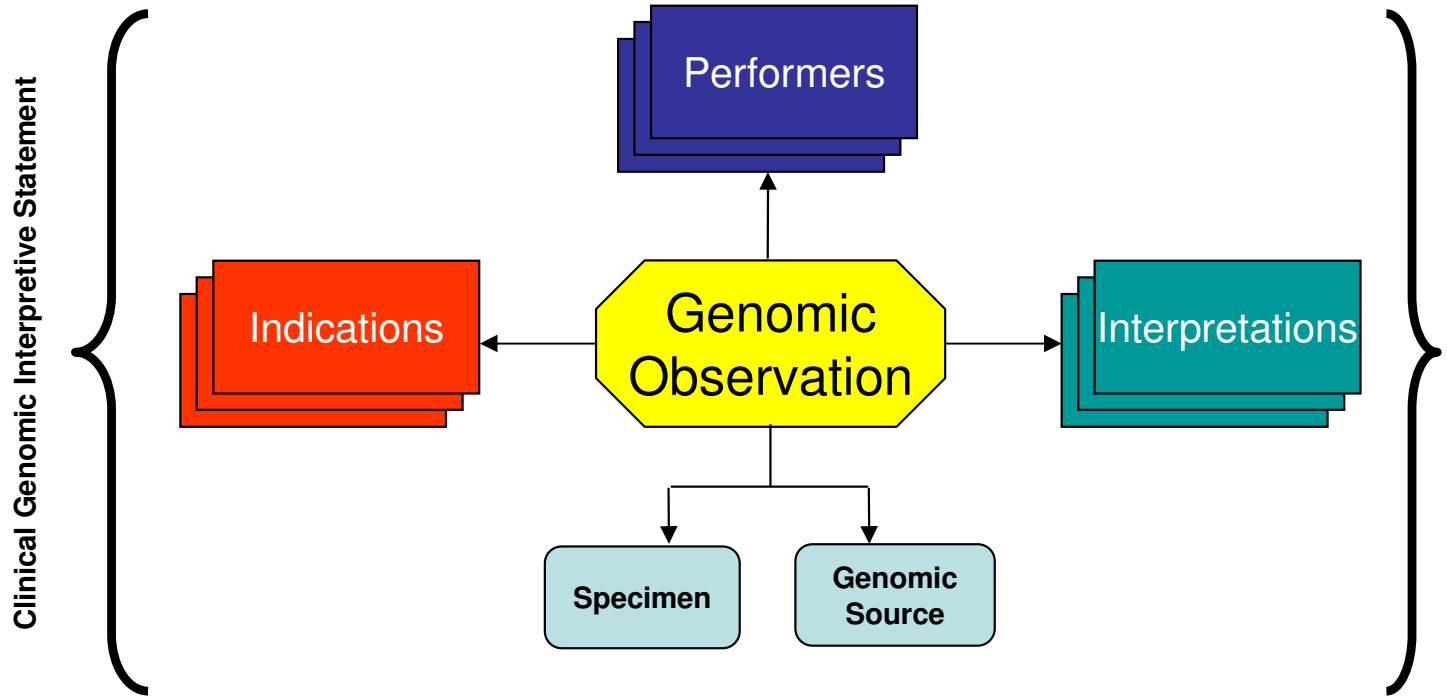
Draft that has not been clinically validated

CDA GTR Section Outline

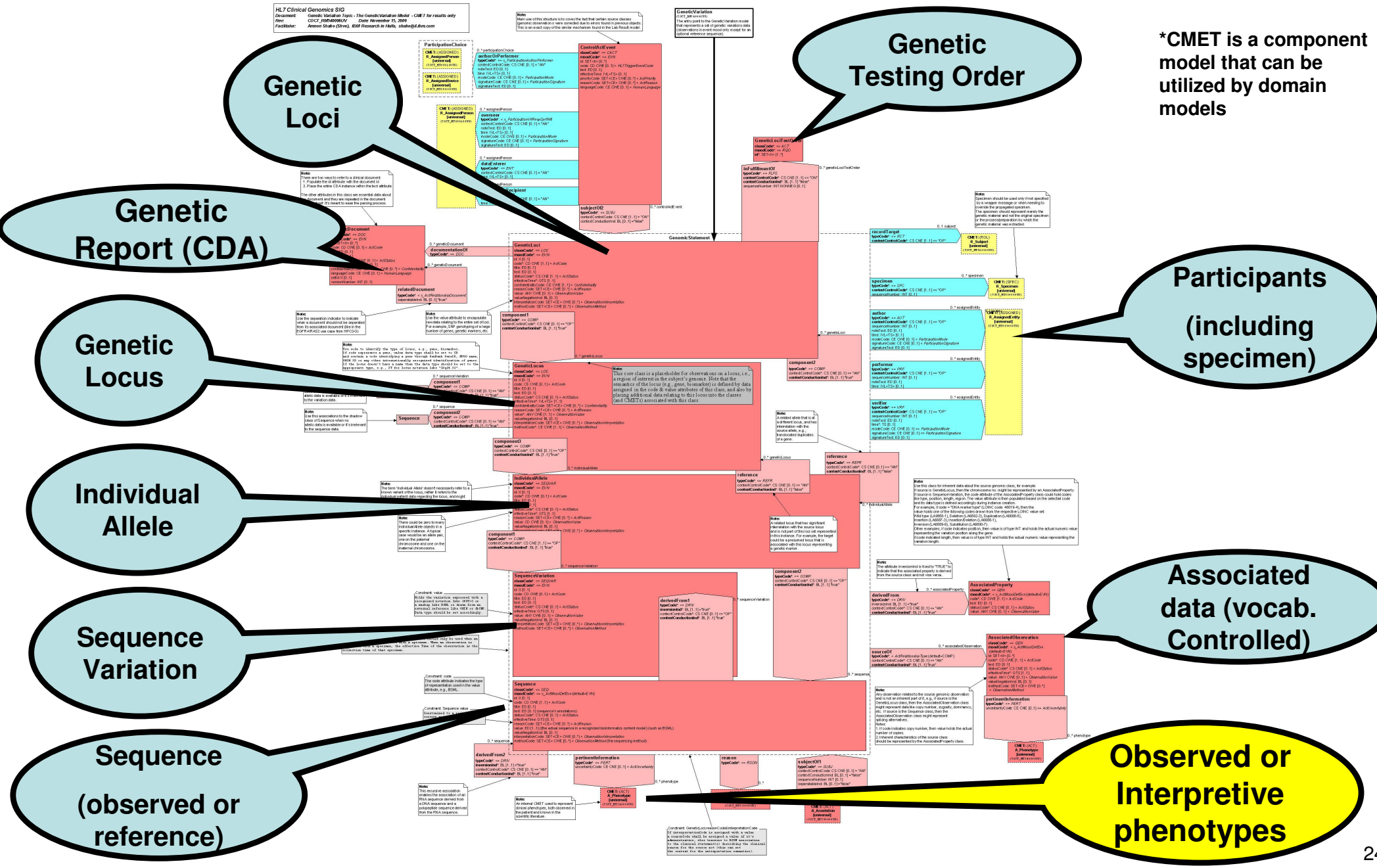


The Clinical Genomic Statement

- An abstract Clinical Genomic Statement (CGS) template that
  - has at its core a genomic observation (e.g., a DNA sequence variation)
  - If it's a major observation, then it should be associated with indications and interpretations, specimen and genomic source class, and optionally associated with performers
  - If it's an associated observation (e.g., amino acid change), then only the genomic observation is populated and optionally the performers
- The CGS abstract template is instantiated by specialized CGS's, e.g., for genetic variations or cytogenetics, as well as for their associated observation

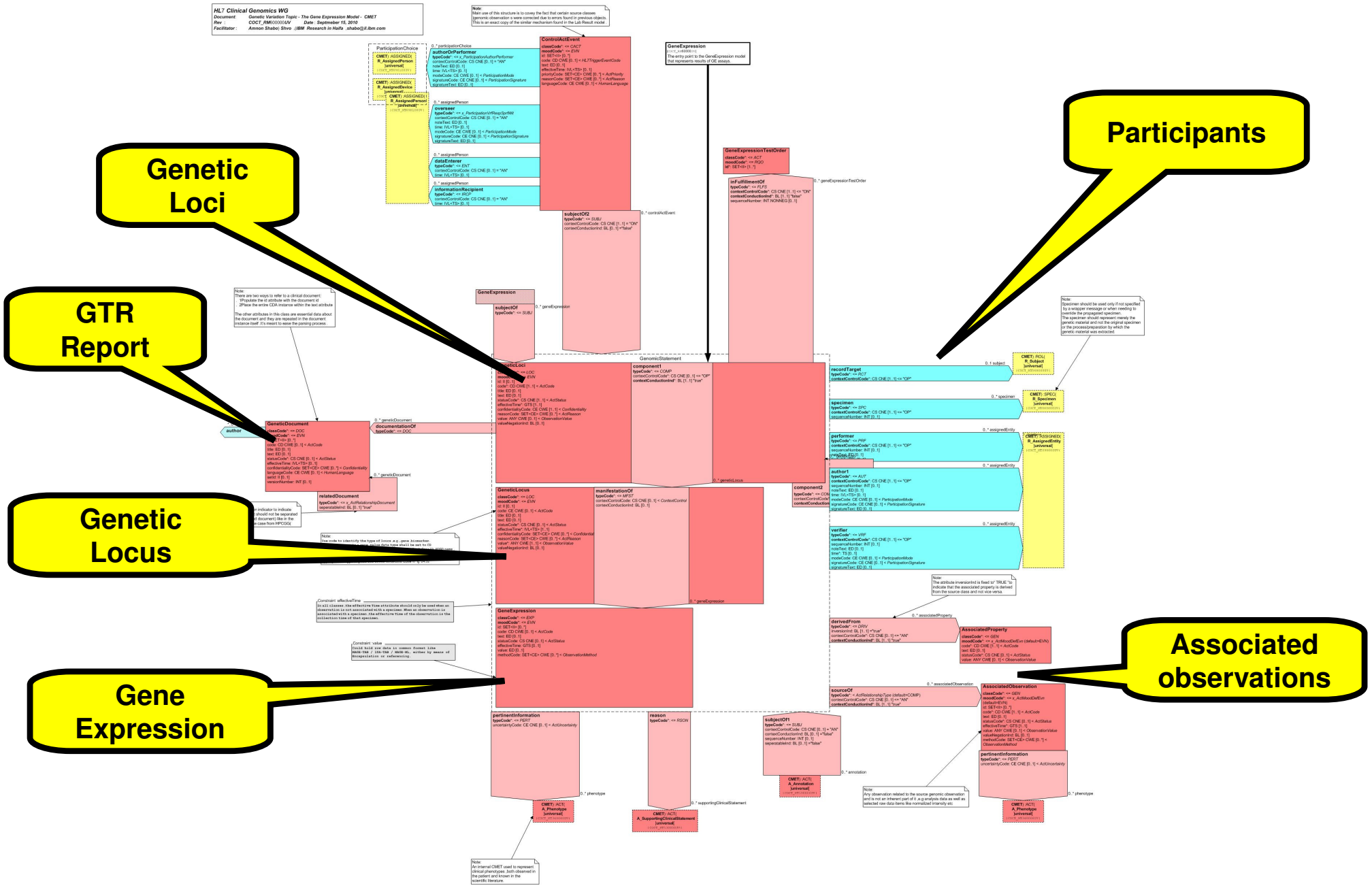


The Genetic Variation CMET* (passed normative in Jan. 2010)



*CMET is a component model that can be utilized by domain models

The Gene Expression CMET Draft



Experimental Implementations

- V3 specs
  - The Genetic Variation and Pedigree models are used in Hypergenes (a European project on essential hypertension, http://www.hypergenes.eu/)
  - The Family History spec is widely used in (e.g., MGH, HHS)
  - The Pedigree and Genetic Variation models are used by the Rizzoli institute in Bologna, Italy for orthopedic genetic diseases
- CDA GTR
  - Used in Korea in the uHealth project (Gil Hospital)
- v2 IG
  - Used by Harvard and Intermountain to send genetic testing results message

Hypergenes – Essential Hypertension Genomics

Challenge & Objectives

- **An EC-FP7 funded project** addressing challenge **HEALTH-2007-2.1.1-2**: Molecular epidemiological studies in existing well characterized European (and/or other) population cohorts
- **Objective**: To define a comprehensive genetic epidemiology disease model of essential hypertension (EH) by integrating new technologies of high-throughput genotyping with sophisticated statistical-mathematical modeling and methods of genetic epidemiology

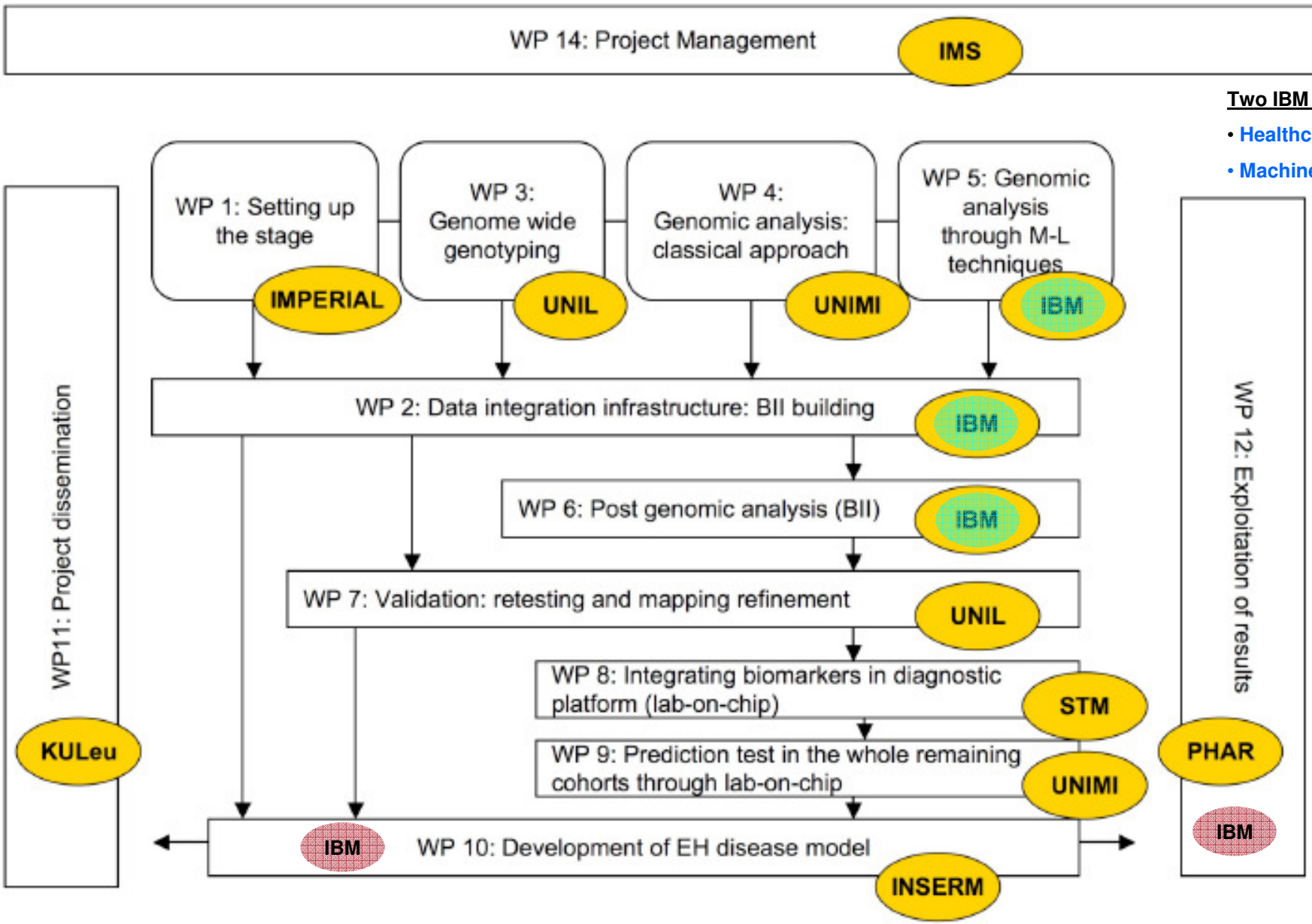
Scientific Coordinator

- State University of Milano

Collaborators

- State University of Milano; Katholieke Universiteit Leuven; Uniwersytet Jagiellonski Collegium Medicum; Sineurra; IMS Research; State Scientific Research Institute of Internal Medicine, Russian Academy of Medical Sciences Siberian Department; Imperial College London; UC San Diego; INSERM - College de France; Warwick Medical School; Prassis-SigmaTau Research Institute, Milano; STMicroelectronics; Losanna & Ginevra University; Pharm-Next; Softeco Sismat Spa, Genoa; Shanghai Institute of Hypertension; Charles University in Prague; Faculty of Medicine in Pilsen; State University of Padova; Medical University of Gdansk

Hypergenes Work Packages



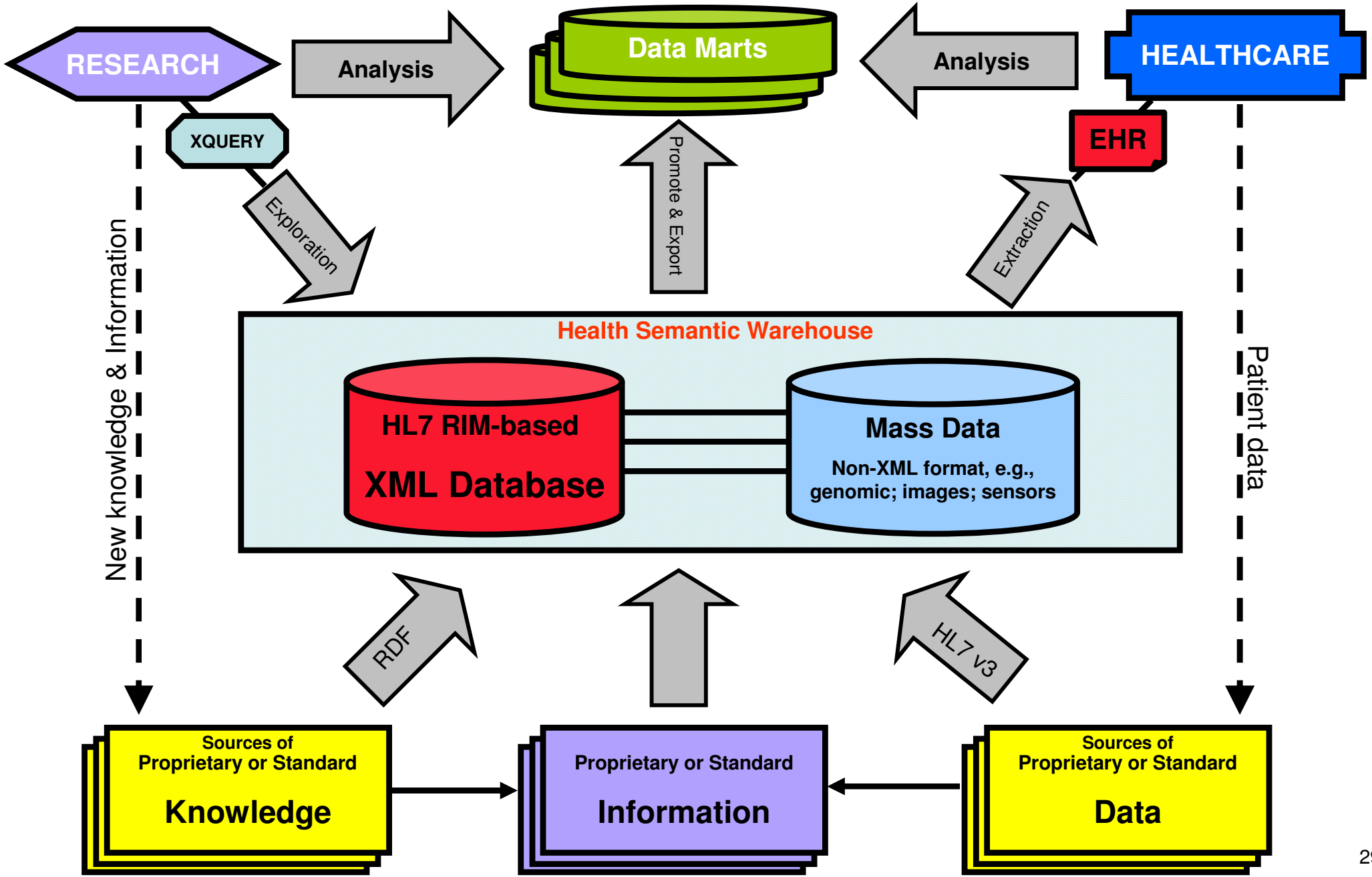
Two IBM Haifa Research Groups:

- Healthcare & Life Sciences Group
- Machine Learning Group

IBM Lead

IBM significant Involvement

The Biomedical Information Infrastructure (BII) Landscape



Information Models over RIMon Warehouse

One instance per subject

CDA Template

Header

subject id

....

Body

reference1 to GV

reference2 to GV

reference2 to PD

....

clinical & environmental observations

GV Template

subject id

Genomic Observations

Phenotypes

Encapsulation or referencing

Raw Genomic Data

subject id

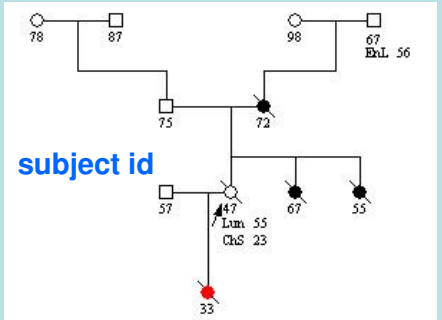
HapMap / BSML / MAGE

or

Relational schemas

optimized for persistency

Pedigree Template



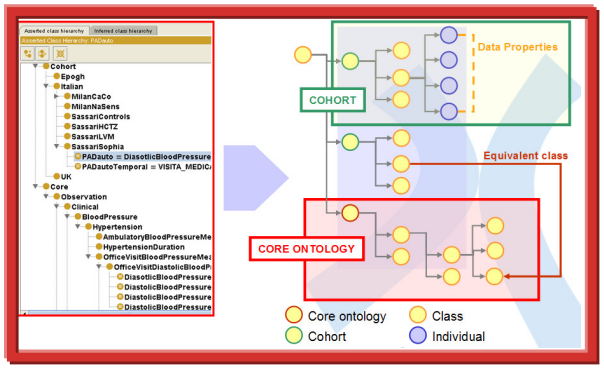
The Hypergenes example

Instance Generation (Data & Knowledge)

Data Source

Adapter

OWL Ontology



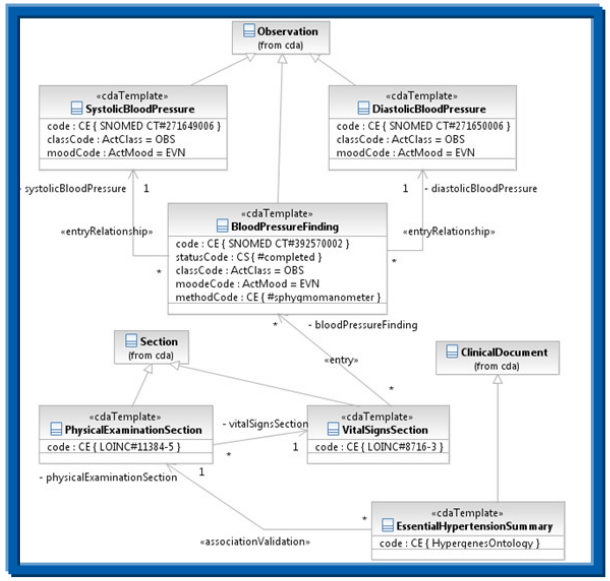
Mapping local Vocabularies

CTS

Instance Generation Engine

Java API

Template Model



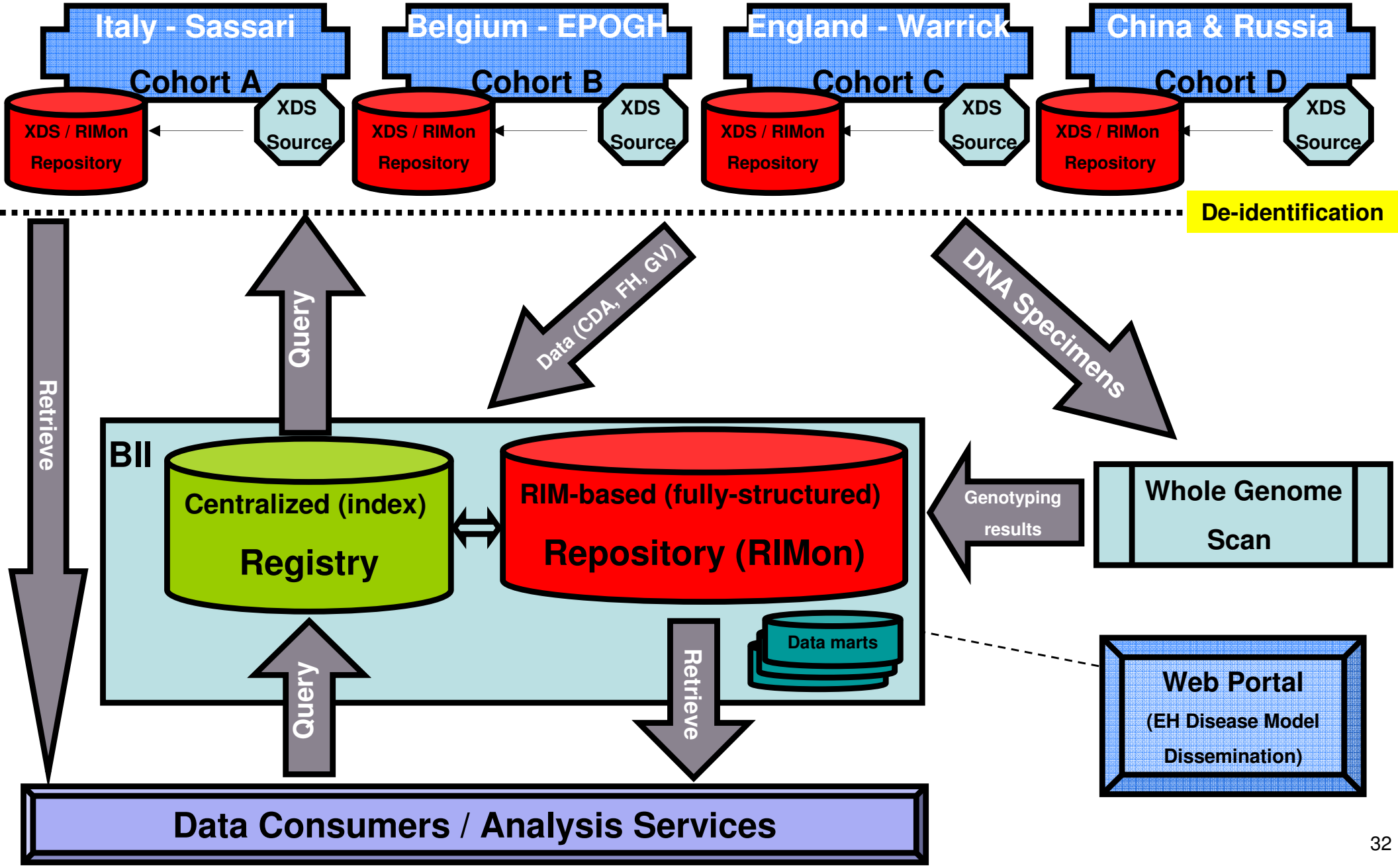
Representing constraints

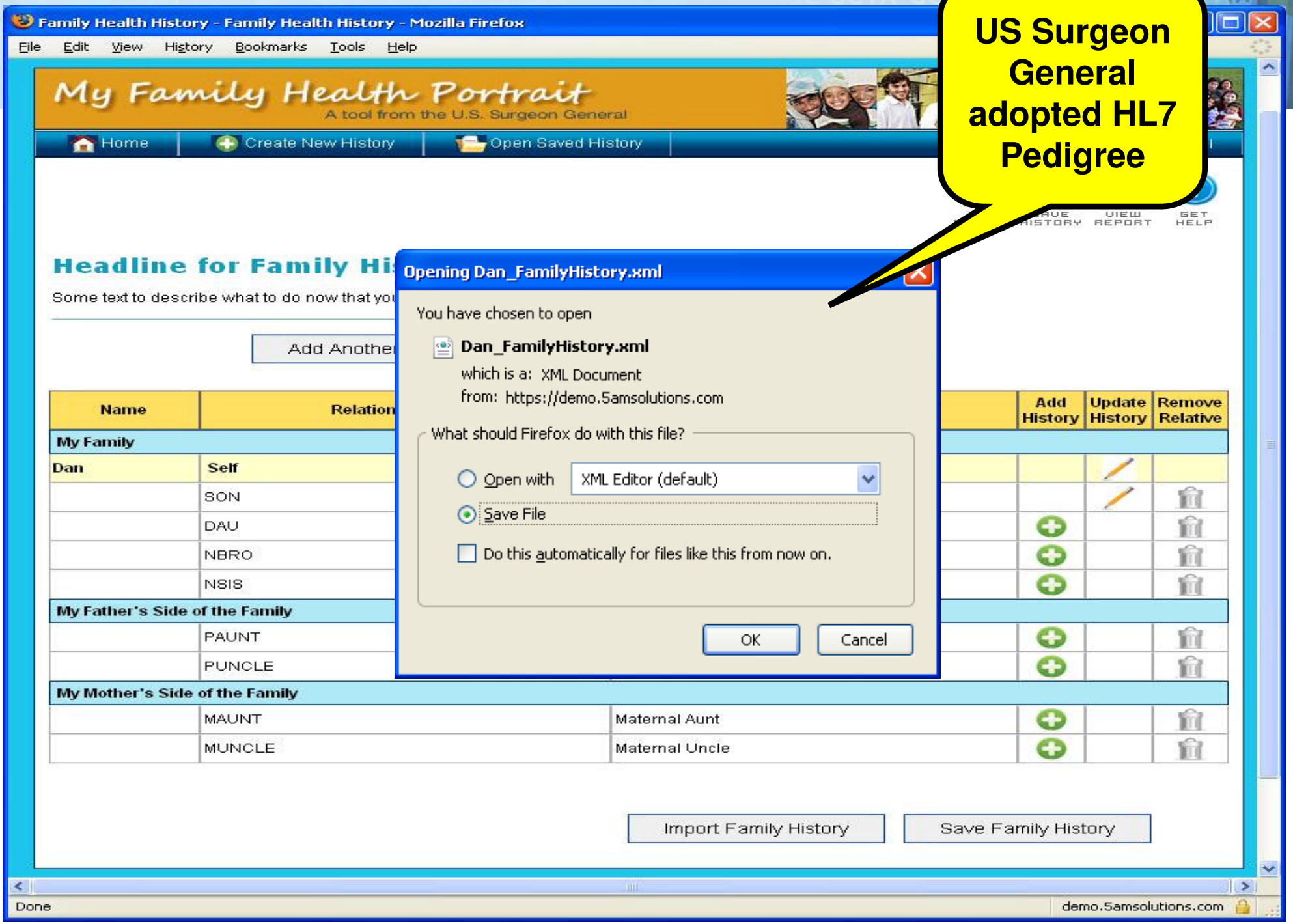
Standard-based Instances

(e.g., CDA)

Conform to the Template Model

Potential Support of Distributed Repositories – Extended IHE XDS



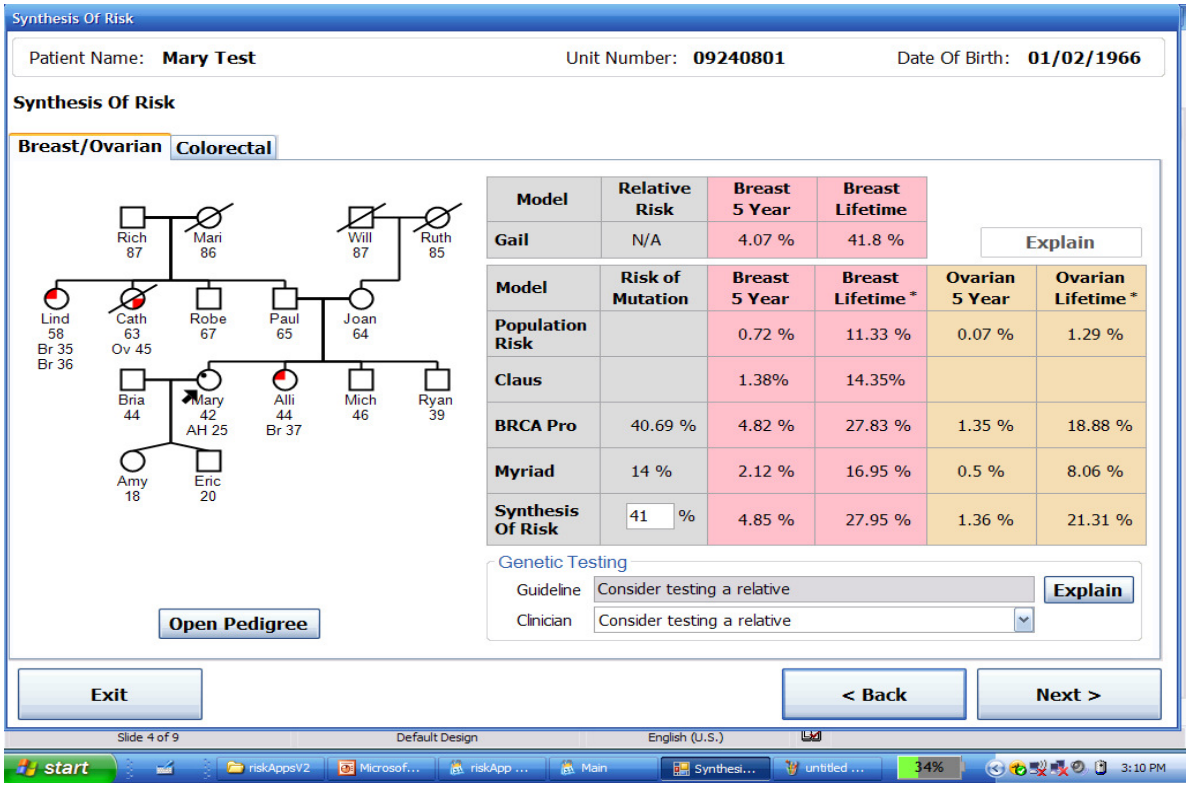


Visit: https://familyhistory.hhs.gov/

33

HughesRiskApps complies with the HL7 standard

- Data can be shared with any HL7 compliant software
- Data can be uploaded or downloaded to any EHR that has a complete family history section and that is HL7 compliant
- http://www.hughesriskapps.net/

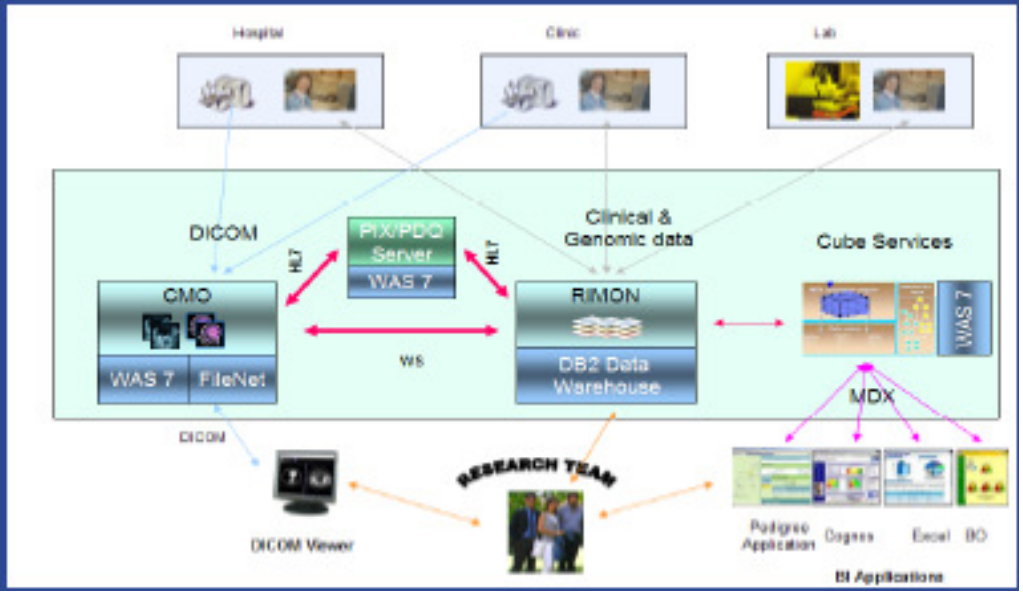


Biomarker Imagining Management

BioMIMS – Rizzoli

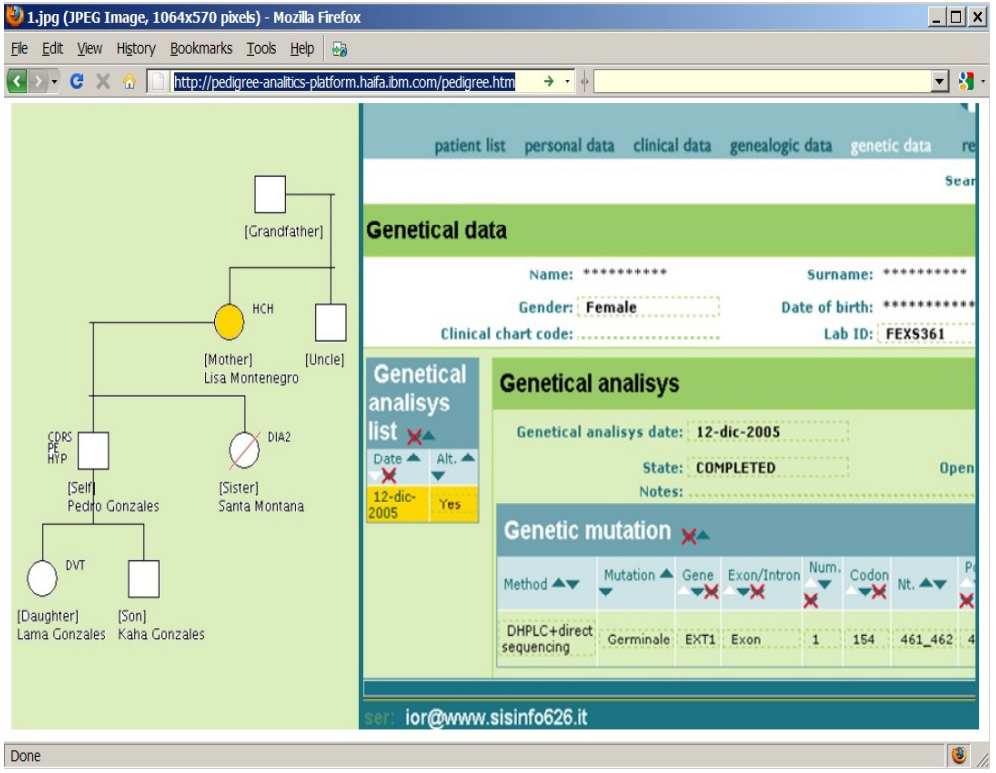
SERVIZIO SANITARIO REGIONALE
EMILIA - ROMAGNA
Istituto Ortopedico Rizzoli di Bologna
Istituto di Ricovero e Cura a Carattere Scientifico

- The Client
  - Medical Genetic Unit of Istituto Ortopecico Rizzoli (IRO)
- Goal
  - Imaging biomarkers hold tremendous potential for accelerating the development of pharmaceuticals and therapeutic devices, as well as for improving the quality of patient care
  - Develop an imaging biomarkers management solution, leveraging the correlation of bio-medical imaging, clinical and genomic data, based on healthcare standards
  - Support sophisticated analytics and queries, with special emphasis on pedigree visualization
- Challenges
  - Collaborative environment with regional, national and international projects, on skeletal genetic diseases
  - Extensive usage of Imaging data, Genomics data, Clinical Data, Pedigree Information



BioMIMS - Pedigree Visualization & Access

- Dynamic pedigree visualization
- Presentation of all available information for the persons in the pedigree
  - **Clinical, genomic data and medical images**
- Standard pedigree representation
  - **HL7 v3 Family History**
  - **Enables standard based pedigree interoperability**
- Enables disease risk assessment



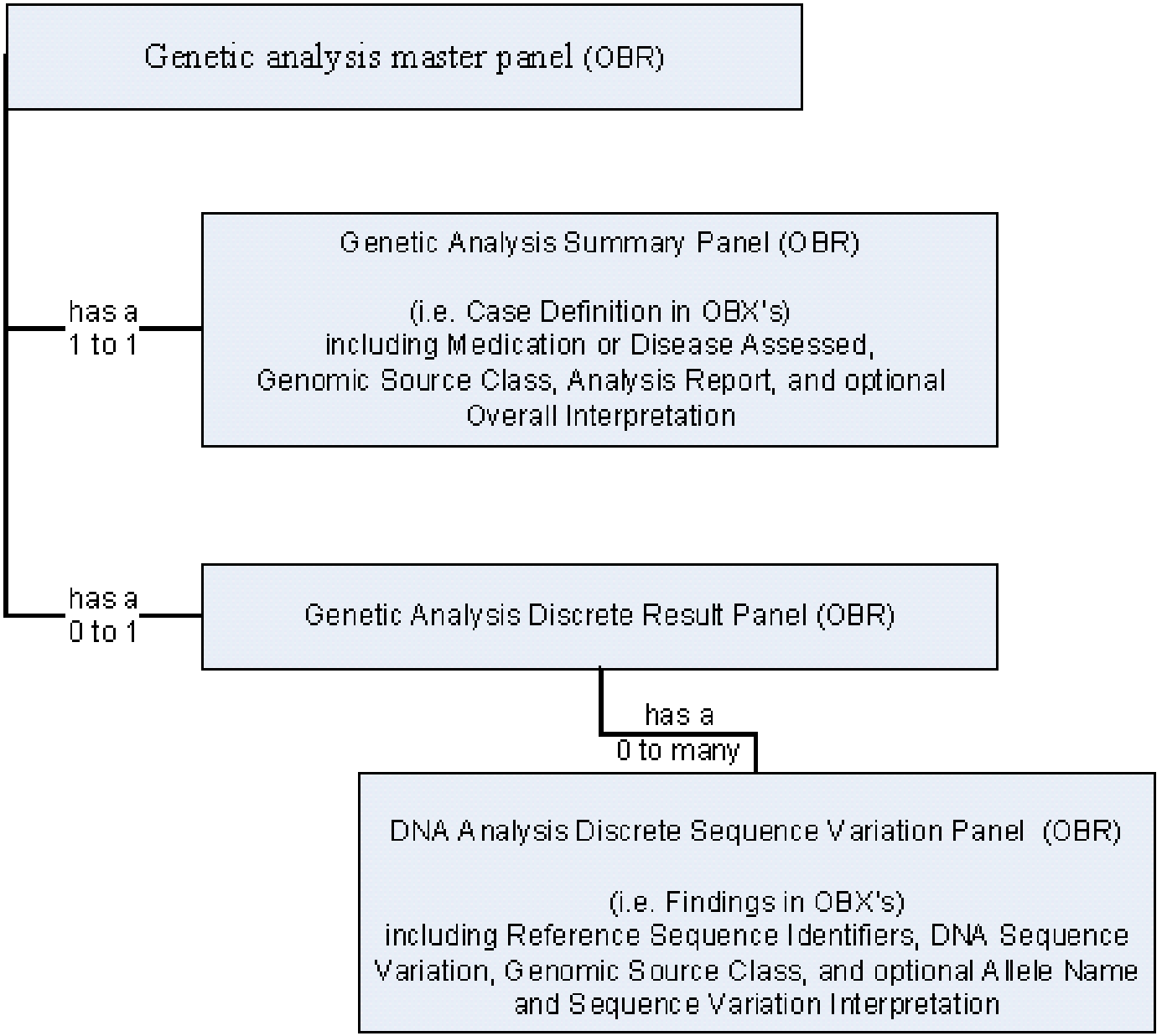
v2 Implementation Guide (of Lab)

- The IG “Genetic Test Result Reporting to EHR” passed informative ballot
- It is modeled after the HL7 Version 2.5.1 Implementation Guide: Orders And Observations; Interoperable Laboratory Result Reporting To EHR (US Realm), Release 1
- Is used in a pilot of information exchange between Partners Healthcare and Intermountain Health Care

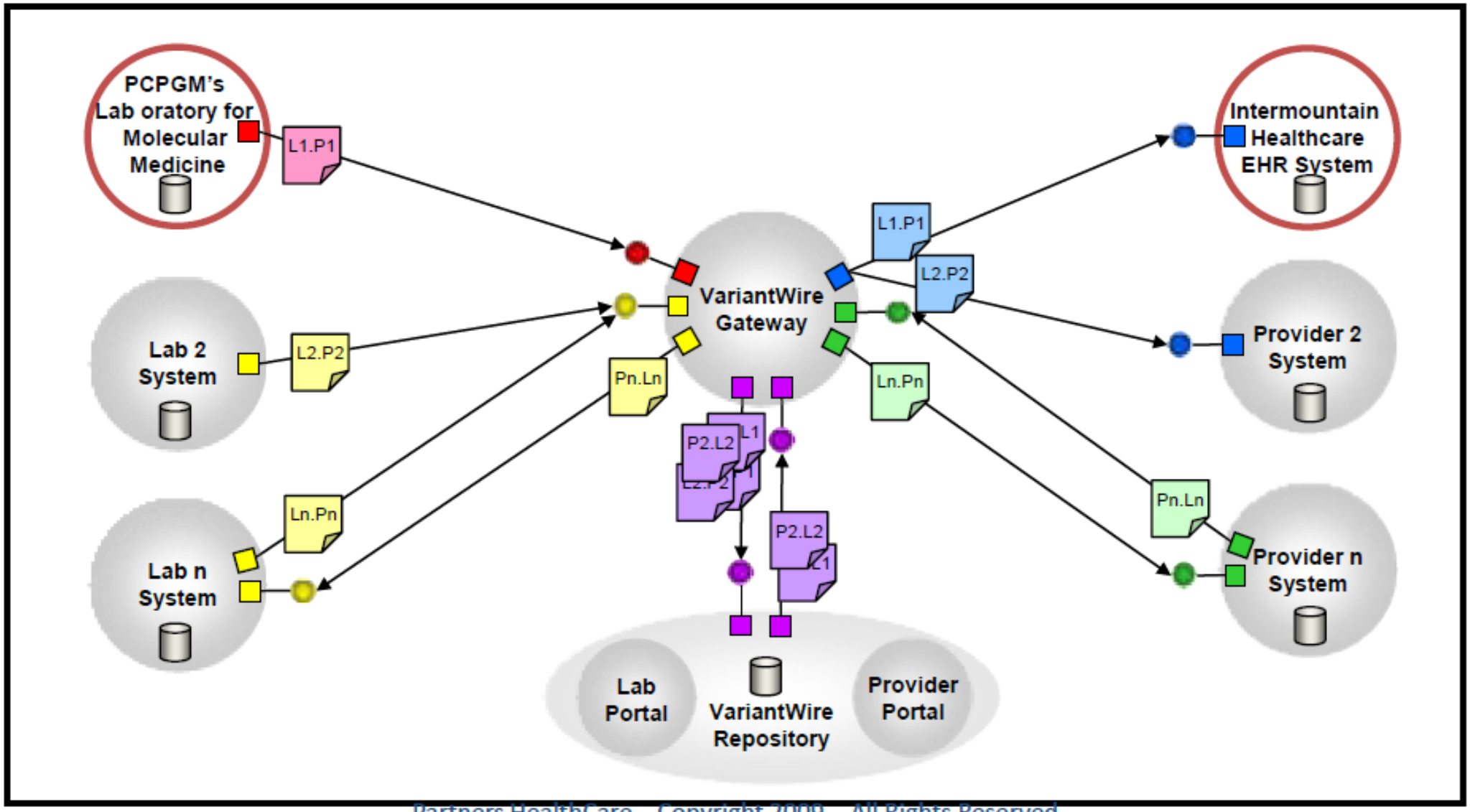
v2 Implementation Guide (of Lab)

- The IG “Genetic Test Result Reporting to EHR” passed informative ballot
- It is modeled after the HL7 Version 2.5.1 Implementation Guide: Orders And Observations; Interoperable Laboratory Result Reporting To EHR (US Realm), Release 1
- Is used in a pilot of information exchange between Partners Healthcare and Intermountain Health Care

The v2 Message Structure



Harvard – Intermountain Exchange Pilot



Source: Emerging Clinical Genomics Standards, Mollie Ullman-Cullere, Oct.15, 2009

v2 Sample Message (Harvard – IHC Pilot)

- OBR|1||PM-08-J00094^HPCGG-LMM^2.16.840.1.113883.3.167.1^ISO|lm_DCM-pnlB_L^Dilated Cardiomyopathy Panel B (5 genes)^99LMM-ORDER-TEST-ID||20080702000000|20080702100909||||||||234567891^Pump^Patrick^^^^^^NPI^L|||||20080703000000|||F||||||00000009^Cardiovascular^99HPCGG-GVIE-INDICATION^^^^^^Clinical Diagnosis and Family History of DCM|&Geneticist&Gene&&&&&NPI^^^^^^^HPCGG-LMM&2.16.840.1.113883.3.167.1&ISO||||||||||||||55233-1^Genetic analysis master panel ^LN

- SPM|1|||119273009&Peripheral blood&SNM3&&&&0707Intl&&Blood, Peripheral||||||||||||20080702000000

- OBR|2||PM-08-J00094-1^HPCGG-LMM^2.16.840.1.113883.3.167.1^ISO|55232-3^Genetic analysis summary panel^LN|||20080702000000||||||||||||||20080703000000|||F||||^PM-08-J00094&HPCGG-LMM&2.16.840.1.113883.3.167.1&ISO

- OBX|1|CWE|51967-8^Genetic disease assessed^LN||399020009^DCM-Dilated Cardiomyopathy^SNM3^^^0707Intl||||||F|20080702100909||||||||||Laboratory for Molecular Medicine^L^22D1005307^^^CLIA&2.16.840.1.113883.4.7&ISO|1000 Laboratory Lane^Ste. 123^Cambridge^MA^99999^USA^B

Summary

- Small group coping with
  - Various HL7 formats: v3, v2 and CDA
  - Clinical & Research environments
- Developing component models (CMET) to be used in other HL7 domains
  - Genetic Variation
  - Gene Expression
- CDA Genetic Testing Report (GTR)
  - Bridge from raw data to human readable reports and bubbled-up data
  - Model-driven development of standards (use of MDHT CDA Editor)
- Call for European Participation…!
  - An out-of-cycle meeting to kick-off European involvement
  - Collocated with major EU venues (e.g., MIE 2011)

The End

- Thank you for your attention… ☺
- Questions? Contact Amnon at shabo@il.ibm.com
- Comments of general interest should be posted to the CG mailing list at clingenomics@lists.hl7.org